

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:16:49 ; Search time 91.75 Seconds

(without alignments)
23.231 Million cell updates/sec

Title: US-09-331-631a-5_COPY_145_210

Perfect score: 375
Sequence: 1 KRDPQREYEDCRHCEQOE.....PQGGSGRYEGEGEKQSDNP 66

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	121	32.3	588	1	VCLB_GOSHI
2	120	32.0	605	1	VCLB_GOSHI
3	78	20.8	763	1	GLH1_CABEL
4	77	20.5	428	1	INVO_CEBAL
5	76	20.3	493	1	INVO_SAGOE
6	75.5	20.1	550	1	BLSA_HUMAN
7	72	19.2	573	1	GLB1_MAIZE
8	72	19.2	911	1	ANDR_PANTR
9	71.5	19.1	1898	1	TRH1_HUMAN
10	71	18.9	544	1	INVO_AOTTR
11	71	18.9	919	1	ANDR_HUMAN
12	71	18.9	1903	1	TAGB_DICDI
13	70	18.7	1023	1	GLT1_DROME
14	70	18.7	1090	1	NIT4_NEOCR
15	69.5	18.5	255	1	LP61_ELITE
16	69	18.4	905	1	SNF5_YEAST
17	69	18.4	966	1	SNF6_YEAST
18	69	18.4	1265	1	CYAS_CANPA
19	68.5	18.3	398	1	PE21_ARATH
20	68.5	18.3	975	1	CDP_CANPA
21	68	18.1	538	1	GAG_MLVEP
22	68	18.1	756	1	KNO1_YEAST
23	68	18.1	2442	1	CBP_HUMAN
24	68	18.1	3828	1	TRX_DROVI
25	67.5	18.0	568	1	INVO_RAF
26	67.5	18.0	807	1	L100_ADE05
27	67	17.9	629	1	PAB2_ARATH
28	67	17.9	646	1	SG1_BOVIN
29	67	17.9	907	1	ANDR_CANFA
30	67	17.9	2476	1	ZAN_PIG
31	66.5	17.7	347	1	INVO_PIG
32	66	17.6	2056	1	CBP1_CABEL
33	65.5	17.5	656	1	DNAI_STRCO

34	65.5	17.5	758	1	YM38_YEAST
35	65.5	17.5	1344	1	IF3A_MOUSE
36	65.5	17.5	1382	1	IF3A_HUMAN
37	65.5	17.5	1462	1	NKCR_HUMAN
38	65	17.3	157	1	HMAA_SCHGR
39	65	17.3	450	1	SRE_DROME
40	65	17.3	693	1	CAUP_DROME
41	64.5	17.2	895	1	ANDR_MACFA
42	64	17.1	467	1	INVO_MOUSE
43	64	17.1	529	1	GAG_MLVDU
44	64	17.1	536	1	GAG_MLVDE
45	64	17.1	539	1	GAG_MLVFS

ALIGNMENTS

RESULT	ID	VCLB_GOSHI	STANDARD	PRT	588 AA
AC	P09801				
DT	01-MAR-1989	(Rel. 10, Created)			
DT	01-MAR-1989	(Rel. 10, Last sequence update)			
DT	13-JUL-1999	(Rel. 38, Last annotation update)			
DE	VICILIN C72 PRECURSOR (ALPHA-GLOBULIN B).				
OS	Gossypium hirsutum (Upland cotton).				
OC	Eukaryota: Viridiplantae: Embryophyta: Tracheophyta: Spermatophyta:				
OC	Magnoliophyta: eudicotyledons: core eudicots: Rosidae, eustosids II;				
OC	Malvales: Malvaceae; Gossypium.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Chlan C.A., Pyle J.B., Legocki A.B., Dure L. III;				
RT	"Developmental biochemistry of cottonseed embryogenesis and				
RT	germination. XVIII. cDNA and amino acid sequences of the members of				
RT	the storage protein families."				
RL	Plant Mol. Biol. 7:475-489(1986).				
CC	-1- FUNCTION: SEED STORAGE PROTEIN.				
CC	-1- SUBCELLULAR LOCATION: CYTOSOL;ER;MEMBRANE-BOUND VACUOLAR PROTEIN				
CC	BODIES.				
CC	-1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,				
CC	CONVICALIN, CONGLACININ, ETC.).				
CC					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC					
DR	EMBL: M16891; AAA33071.1; -.				
DR	PIR: A30838; FMCNAB.				
DR	HSSP: P50477; ICAX.				
DR	INTERPRO: IPR001113; -.				
DR	PFAM: PF00546; Seedstore_7s; 1.				
KW	Seed storage protein; Signal.				
FT	SIGNAL	1	25		
FT	CHAIN	26	588		
SQ	SEQUENCE	588 AA; 69729 MW; 638699829AB8ADEB CRC64;			

Query Match 32.3%; Score 121; DB 1; Length 588;

Best local Similarity 34.3%; Pred. No. 2.7e-05;

Matches 35; Conservative 11; Mismatches 16; Indels 40; Gaps 6;

QY	3	DPQOREYEDCRHCEQOEPRLOYOCQORC-----QEQO-----RQHGREGDLMNPQ 48
DB	82	DPQRR-YEECOECROOEERPOQOCORCLKREOEQOQOQOFQOEQOHCHOQE-QRPE 139
QY	49	RGS-----GRY-----FEGEKQSDNP 66
DB	140	RKOOCVRECRERYOENPWRREDEAEDEETEEGDEQSHNP 181

```

RESULT 2
VCLA_GOSHI STANDARD; PRT; 605 AA.
AC P09799;
DT 01-MAR-1989 (rel. 10, Created)
DT 01-MAR-1989 (rel. 10, Last sequence update)
DT 15-JUL-1999 (rel. 38, Last annotation update)
DE VICILIN GC72-A PRECURSOR (ALPHA-GLOBULIN A).
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Gossypium.
RN (1)
RP SEQUENCE FROM N.A.
RA Chlan C.A., Borroto K., Kamalay J.A., Dure L. III;
RT "Developmental biochemistry of cottonseed embryogenesis and
RT germination. XIX. Sequences and genomic organization of the alpha
RT globulin (vicilin) genes of cottonseed."
RL Plant Mol. Biol. 9:533-546(1987).
CC -1- FUNCTION: SEED STORAGE PROTEIN.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC MEMBRANE-BOUND VACUOLAR PROTEIN
CC BODIES.
CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICTILIN, CONGLICININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M19378; AAA33069.1; -.
DR PIR: S06398; S06398.
DR HSSP: P50477; 1CAK.
DR INTERPRO: IPR001113; -.
DR PFAM: PF00546; Seedstore_7s; 1.
KW Seed storage protein; Signal.
FT SIGNAL 1 23
FT CHAIN 24 605 VICILIN GC72-A.
FT SEQUENCE 605 AA; 71049 MW; C9DB9371C976953B CRC64;

Query Match 32.0%; Score 120; DB 1; Length 605;
Best Local Similarity 31.4%; Pred. No. 3.5e-05;
Matches 33; Conservative 14; Mismatches 16; Indels 42; Gaps 5;

QY 3 DPQGRREDCRRHCEQOEPRL-----QY-----QCQRQCEQORONG 39
DB 79 DPQRR-YQDCRQHCQOEERLRPHCEQSCREYEQOQOQDPKQKRECOQRCQMOEORPE 137
QY 40 R-----GCDLNPQRQSGRYE-----EGEEKOSDNP 66
DB 138 RQQCVKCEQREYQEDPMKGERENKWRKEEERESDGEQOQRNRP 182

RESULT 3
GLH1_CAEEL STANDARD; PRT; 763 AA.
AC P34689; Q9TXH4;
DT 01-FEB-1994 (rel. 28, Created)
DT 01-OCT-2000 (rel. 40, Last sequence update)
DT 01-OCT-2000 (rel. 40, Last annotation update)
DE ATP-DEPENDENT RNA HELICASE GLH-1 (GERMLINE HELICASE-1).
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RT STRAIN-BRISTOL N2;

```

```

RX MEDLINE-94022363; PubMed-8415696;
RA Roussel D.L., Bennett K.L.;
RT "glh-1, a germ-line putative RNA helicase from Caenorhabditis, has
RT four zinc fingers."
RL Proc. Natl. Acad. Sci. U.S.A. 90:9300-9304(1993).
RN [2]
RP REVISIONS TO 83-138; 275; 288 AND 398.
RA Roussel D.L., McCrone J.S., Smith P.A., Gruidl M.E., Bennett K.L.;
RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: PROBABLE ATP-BINDING RNA HELICASE.
CC -1- DEVELOPMENTAL STAGE: DURING GERM-LINE PROLIFERATION.
CC -1- SIMILARITY: TO OTHER "DEAD" BOX FAMILY HELICASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: L19948; AAC27384.1; -.
DR PIR: A48686; A48686.
DR HSSP: P05888; 1AAR.
DR INTERPRO: IPR000629; -.
DR INTERPRO: IPR001410; -.
DR INTERPRO: IPR001650; -.
DR INTERPRO: IPR001878; -.
DR PFAM: PF00270; DEAD; 1.
DR PFAM: PF00271; helicase_C; 1.
DR PFAM: PF00098; zf-CCHC; 4.
DR PRINTS: PR00939; C2HCZNFINGER.
KW Helicase; ATP-binding; RNA-binding; Zinc-finger; Repeat.
DR PROSITE: PS00039; DEAD_ATP_HELICASE; 1.
FT DOMAIN 24 93
FT REPEAT 24 33
FT REPEAT 34 43
FT REPEAT 44 53
FT REPEAT 54 63
FT REPEAT 64 73
FT REPEAT 74 83
FT REPEAT 84 93
FT DOMAIN 207 236 GLY-RICH.
FT ZN_FING 160 173 C2HC-TYPE.
FT ZN_FING 185 198 C2HC-TYPE.
FT ZN_FING 244 257 C2HC-TYPE.
FT ZN_FING 264 277 C2HC-TYPE.
FT NP_BIND 385 392 ATP (BY SIMILARITY).
FT SITE 489 502 DEAD BOX.
FT SEQUENCE 763 AA; 79792 MW; ADB69DE286A028D6 CRC64;

Query Match 20.8%; Score 78; DB 1; Length 763;
Best Local Similarity 28.6%; Pred. No. 0.7;
Matches 22; Conservative 11; Mismatches 26; Indels 18; Gaps 3;

QY 2 RDPQREYEDCRRHCEQOEPRLOYQO-----RRCQEQOROH-----GRGDL 44
DB 164 QDPGRHS-SDCEPRKREPRVCYNQCQPGHNSRCEERKPREGRTGGFGGAGCGNNG 222
QY 45 MNPQRCGSGRYEGEEK 61
DB 223 GNDGFGDGFGGGEER 239

RESULT 4
INVO_CEBAL STANDARD; PRT; 428 AA.
AC P24709;
DT 01-MAR-1992 (rel. 21, Created)
DT 01-MAR-1992 (rel. 21, Last sequence update)
DT 01-FEB-1996 (rel. 33, Last annotation update)
DE INVOLUCRIN.

```

GN	IVL.
OS	Cebus albifrons (White-fronted capuchin).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae; Cebus.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=OEOSOPHAGEAL FIBROBLAST;
RC	MEDLINE=92114750; PubMed=1766560;
RX	Phillips M., Rice R.H., Djian P., Green H.,
RA	"The involucrin genes of the white-fronted capuchin and cottontop
RT	tamarin: the plathyrrhine middle region.";
RL	Mol. Biol. Evol. 8:579-591(1991).
CC	-I- FUNCTION: INVOLUCRIN IS A KERATINOCYTE PROTEIN THAT FIRST APPEARS
CC	IN THE CELL CYTOSOL, BUT ULTIMATELY BECOMES CROSS-LINKED TO
CC	MEMBRANE PROTEINS BY TRANSGLUTAMINASE. ALL THAT RESULTS IN THE
CC	FORMATION OF AN INSOLUBLE ENVELOPE BENEATH THE PLASMA MEMBRANE.
CC	-I- TISSUE SPECIFICITY: PRESENT IN KERATINOCYTES OF EPIDERMIS AND
CC	OTHER STRATIFIED SQUAMOUS EPITHELIA.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; M67478; AAA35405.1; -.
DR	INTERPRO; IPR000354; -.
DR	INTERPRO; IPR002360; -.
DR	PFAM; PF00904; Involucrin; 28.
DR	PROSITE; PS00795; INVOLUCRIN; 1.
DR	Keratinocyte, Repeat.
SO	SEQUENCE 428 AA; 50126 MW; 03AB7A15FC34EB02 CRC64;

```

Query March 20.54; Score 77; DB 1, Length 428;
Best Local Similarity 29.24; Pred. NO. 0.52;
Matches 21; Conservative 17; Mismatches 18; Indels 16; Gaps

QY      8 EYEDRRHCEQEEQRLQY-----QCRCRCEQQRHQRGSGDLNMP-ORGSGR 54
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      257 QQEQQLKHLHEQEQLEHLHEQEBELKHLDEQEQLEHLHEQEQ---LELPQEQGSK 313
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      55 YEEGEEKQSDNP 66
      : : : : : : : :
Db      314 HLEQEEKQLHNP 325

RESULT 5
INVO_SAGE
ID      INVO_SAGE      STANDARD;      PRT;      493 AA.
AC      P24712;
DT      01-MAR-1992 (Rel. 21, Created)
DT      01-MAR-1992 (Rel. 21, Last sequence update)
DT      01-FEB-1996 (Rel. 33, Last annotation update)
DE      INVOLUCRIN.
GN      IVL.
OS      Sagalinus oedipus (Cotton-top tamarin).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Sagalinus.
      [1]
      SEQUENCE FROM N.A.
RP      TISSUE-VAGINAL_FIBROBLAST;
RC      MEDLINE=92114750; PubMed=1766360;
RA      Phillips M., Rice R.H., Djian P., Green H.;
RT      "The involucrin genes of the white-fronted capuchin and cottontop
      tamarin: the platyrrhine middle region.";
RL      Mol. Biol. Evol. 8:579-591(1991).
CC      -I- FUNCTION: INVOLUCRIN IS A KERATINOCTE PROTEIN THAT FIRST APPEARS
      IN THE CELL CYTOSOL, BUT ULTIMATELY BECOMES CROSS-LINKED TO
      MEMBRANE PROTEINS BY TRANSGLUTAMINASE. ALL THAT RESULTS IN THE
      FORMATION OF AN INSOLUBLE ENVELOPE BENEATH THE PLASMA MEMBRANE.
CC

```

CC -1- TISSUESPECIFICITY: PRESENT IN KERATINOCYTES OF EPIDERMIS AND
CC OTHER STRATIFIED SQUAMOUS EPITHELIA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M67477; AAA36950.1; -.
DR HSSP: P80220; IDIP.
DR INTERPRO: IPRO00354; -.
DR INTERPRO: IPRO02360; -.
DR PfAM: PF00904; Involucrin:34.
DR PROSITE: PS00795; INVOLUCRIN: 1.
KW Keratinoctyte. Repeat.
SQ SEQUENCE 493 AA; 57920 MW; A626FCAE7799AEC9 CRC64;

Query Match	20.3%;	Score 76;	DB 1;	Length 493;
Best Local Similarity	36.6%;	Pred. No. 0.75;		
Matches	26;	Conservative 11;	Mismatches 20;	Indels 14;
				Gaps
OY	8	EYECRRGRCGEOERLTOYQOORCO	-----EQO-----	QNGKSGDLMP-QNGSGGRY 55
Db	316	FOEQLNLNLEFOEQLNLNLEQOEGOLENLEHLEQOGLKNLEOH	-----EGOLELPEQOVLGOSKH 373	
OY	56	EEGEGKQSDNP	66	
Db	374	LEOEKOLENP	384	

	RESULT	6			
ID	BLSA_HUMAN		STANDARD:	PRT:	550 AA.
AC	002832;				
DT	01-JUN-1994 (Rel. 29, Created)				
DT	01-JUN-1994 (Rel. 29, Last sequence update)				
DT	01-OCT-1994 (Rel. 30, Last annotation update)				
DE	B-LYMPHOCYTE ANTIGEN PRECURSOR (B-LYMPHOCYTE SURFACE ANTIGEN) (721P).				
OS	Homo sapiens (human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=PLACENTA.				
RX	MEDLINE=93066251; PubMed=1438229;				
RA	Voland J.R., Wyzynkowski R.J., Huang M., Dutton R.W.;				
RT	"Cloning and sequencing of a trophoblast-endothelial-activated				
RT	lymphocyte surface protein: cdna sequence and genomic structure.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 89:10425-10429(1992).				
CC	-I- FUNCTION: THIS PROTEIN MAY BE INVOLVED IN B-CELL ACTIVATION.				
CC	MAY ALSO BE INVOLVED IN SIGNAL TRANSDUCTION AND GENE REGULATION.				
CC	-I- SUBUNIT: MONOMER.				
CC	-I- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED (PROBABLY).				
CC	-I- TISSUE SPECIFICITY: EXPRESSED IN ACTIVATED LYMPHOCYTES, MOST				
CC	VASCULAR ENDOTHELIIUM, AND SYNCTIOROPOBLAST.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
EMBL:	M939578; AAA36187.1; -				
DR	PIR; A46419.				
DR	Glycoprotein; Signal; Membrane.				
ET	SIGNAL 1 21				
ET	CHAIN 22 550				
ET	B-LYMPHOCYTE ANTIGEN.				

```

FT DOMAIN 344 365 LEUCINE-ZIPPER.
FT CAROHD 85 85 N-LINKED (GLCMAC. . .) (POTENTIAL).
FT CAROHD 138 138 N-LINKED (GLCMAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA: 64120 MW: 02989338A5FB80F7 CRC64:

Query Match
Best Local Similarity 34.4%; Score 75.5; DB 1; Length 550;
Matches 21; Conservative 7; Mismatches 16; Indels 17; Gaps 3;

OY 4 POOREVEDCRH-----CEQOEPRLOYOCRCQROHGRGDLMPORG 51
DB 470 PADRVYRLCERHHAAPPPAPRCPOGEPG-PPRCGRSRGKREGRGKG----PMGG 524

OY 52 S 52
DB 525 S 525

RESULT 7
GLB1_MAIZE STANDARD; PRT; 573 AA.
AC P15590;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GLOBULIN-1 S ALLELE PRECURSOR (GLB1-S) (7S-LIKE).
GN GLB1.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=CV. INBERD LINE VA26;
RA Belanger F.C., Kriz A.L.;
RT "Molecular characterization of the major maize embryo globulin encoded
  by the Glb1 gene."
  Plant Physiol. 91:636-643(1989).
  [2]
RN [12]
RP SEQUENCE OF 87-100.
RX MEDLINE=89374022; PubMed=2775172;
RA Kriz A.L.;
RT "Characterization of embryo globulins encoded by the maize Glb
  genes."
  Biochem. Genet. 27:239-251(1989).
CC -I- PFM: THREE PROTEIN-PROCESSING STEPS OCCUR IN THE FORMATION OF THE
  MAJURE PROTEIN FROM THE PRIMARY TRANSLATION PRODUCT.
CC -I- POLYMORPHISM: THE THREE MOST COMMONLY OCCURRING GLB1 ALLELES HAVE
  THE DESIGNATION L, I, AND S FOR LARGE, INTERMEDIATE, AND SMALL
  PROTEINS, RESPECTIVELY.
CC -I- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
  CONVICILIN, CONGLYCININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M24845; AAA33467.1; -.
DR HSSP: P50477; ICAM.
DR MAIZEDB: 30181; -.
DR INTERPRO: IPR001113; -.
DR PFAM: PF00546; Seedstore_7s; 1.
KW Seed storage protein; Signal.
FT SIGNAL 1 18 OR 21 (POTENTIAL).
FT PROPEP 19 86
FT CHAIN 87 573 GLOBULIN-1 S ALLELE.
FT CAROHD 349 349 N-LINKED (GLCMAC. . .) (POTENTIAL).
SQ SEQUENCE 573 AA: 65029 MW: 525EDID00A062976 CRC64:

```

```

Query Match.
Best Local Similarity 34.8%; Score 72; DB 1; Length 573;
Matches 23; Conservative 4; Mismatches 21; Indels 18; Gaps 4;

OY 12 CRHCE---QOEPRLOYC---ORCQEOORHGRGDLMPORGSG-----RYEG 58
DB 40 CYRRCEDRPWHQPRCLCEQCREEREKROERSHREAD-----RSGEGSSEDERERECK 94

OY 59 EKKQSD 64
DB 95 EKKQSD 100

RESULT 8
ANDR_PANTR
ID ANDR_PANTR STANDARD; PRT; 911 AA.
AC O97775;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ANDROGEN RECEPTOR (DIHYDROTESTOSTERONE RECEPTOR).
GN AR OR NR3C4.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98404153; PubMed=9732460;
RA Choong C.S., Kempainen J.A., Wilson E.M.;
RT "Evolution of the primate androgen receptor: a structural basis for
  disease."
  J. Mol. Evol. 47:334-342(1998).
CC -I- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
  THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
  PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC -I- SUBCELLULAR LOCATION: NUCLEAR.
CC -I- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
  A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -I- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC NR3 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U94177; AAC73048.1; -.
DR HSSP: P06536; IRGD.
DR INTERPRO: IPR000536; -.
DR INTERPRO: IPR001103; -.
DR INTERPRO: IPR001628; -.
DR PFAM: PF00104; hormone_rec; 1.
DR PFAM: PF00105; zf-C4; 1.
DR PRINTS: PR00047; STEROIDFINGER.
DR PRINTS: PR00521; ANDROGENR.
DR PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
  Zinc-finger; Steroid-binding.
FT DOMAIN 1 549
FT DNA_BIND 551 616 MODULATING (BY SIMILARITY).
FT ZN_FING 551 571 C4-TYPE ZINC FINGERS (TWO).
FT ZN_FING 587 611 C4-TYPE.
FT DOMAIN 682 911 LIGAND-BINDING.
FT DOMAIN 57 78 POLY-GLN.
FT DOMAIN 84 88 POLY-GLN.
FT DOMAIN 192 196 POLY-GLN.
FT DOMAIN 371 380 POLY-PRO.
FT DOMAIN 395 401 POLY-ALA.
FT DOMAIN 448 464 POLY-GLY.

```

[illegible]

```
CC -I- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-stb.ch/announce/
CC or send an email to license@isb-stb.ch).
-----
DR PIR_109190; AAA65582.1; -.
DR PIR_A45973; A45973.
DR HSP_P02633; IBOC.
DR MIM_190370; -.
DR INTERPRO: IPR001751; -.
DR INTERPRO: IPR002048; -.
DR PFAM: PF01023; S_100; 1.
DR PFAM: PF00036; ehand; 1.
DR PROSITE: PS00018; EF_HAND; 1.
DR PROSITE: PS00303; S100_CABP; 1.
KW Repeat; Calcium-binding.
FT DOMAIN 1 91
FT CA_BIND 22 33 SITE I (LOW AFFINITY) (POTENTIAL).
FT CA_BIND 62 73 SITE II (HIGH AFFINITY) (POTENTIAL).
FT DOMAIN 314 390 R-R-E-Q-E-E-R-R-E-Q-Q-L.
FT REPEAT 314 326 1-1 (APPROXIMATE).
FT REPEAT 327 339 1-2 (APPROXIMATE).
FT REPEAT 340 351 1-3 (APPROXIMATE).
FT REPEAT 352 364 1-4.
FT REPEAT 365 377 1-5.
FT REPEAT 378 390 1-6.
FT DOMAIN 391 444 9 X 6 AA TANDEM REPEATS OF R-R-E-Q-Q-L.
FT REPEAT 391 396 2-1.
FT REPEAT 402 402 2-2.
FT REPEAT 403 408 2-3.
FT REPEAT 409 414 2-4.
FT REPEAT 415 420 2-5.
FT REPEAT 421 426 2-6.
FT REPEAT 427 432 2-7.
FT REPEAT 433 438 2-8.
FT REPEAT 439 444 2-9.
FT DOMAIN 444 702 9 X 28 AA APPROXIMATE TANDEM REPEATS.
FT REPEAT 923 1162 8 X 30 AA TANDEM REPEATS.
FT REPEAT 923 952 4-1.
FT REPEAT 953 982 4-2.
FT REPEAT 983 1012 4-3.
FT REPEAT 1013 1042 4-4.
FT REPEAT 1043 1072 4-5.
FT REPEAT 1073 1102 4-6.
FT REPEAT 1103 1132 4-7.
FT REPEAT 1133 1162 4-8.
FT DOMAIN 1250 1849 23 X 26 AA APPROXIMATE TANDEM REPEATS.
FT CONFLICT 1752 1752 F-> L (IN REF. 2).
FT CONFLICT 1794 1801 OEPRQRYR-> RSFTGSGTG (IN REF. 2).
FT CONFLICT 1857 1857 O-> K (IN REF. 2).
FT CONFLICT 1880 1880 V-> G (IN REF. 2).
SQ SEQUENCE 1898 AA; 247219 MM; A7AB5947FB62E31D CRC64;
```

```

RESULT 10
INVO_AOTTR STANDARD: PRT: 544 AA.
AC P24708:
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE INVOLUCRIN.
GN IVL.
OS Aotus trivirgatus (Night monkey) (Douroucoulli).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Aotinae; Aotus.
RN [1]
RP MEDLINE=90014142; PubMed=2507864;
RA "The involucrin gene of the owl monkey: origin of the early region.";
RL Mol. Biol. Evol. 6:460-468(1989).
CC -I- FUNCTION: INVOLUCRIN IS A KERATINOCYTE PROTEIN THAT FIRST APPEARS
CC IN THE CELL CYTOSOL, BUT ULTIMATELY BECOMES CROSS-LINKED TO
CC MEMBRANE PROTEINS BY TRANSGLUTAMINASE. ALL THAT RESULTS IN THE
CC FORMATION OF AN INSOLUBLE ENVELOPE BENEATH THE PLASMA MEMBRANE.
CC -I- TISSUE SPECIFICITY: PRESENT IN KERATINOCYTES OF EPIDERMIS AND
CC OTHER STRATIFIED SQUAMOUS EPITHELIA.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M25313; AAA35375.1; -
DR HSSP: P80220; IDIP.
DR INTERPRO: IPR000354; -
DR INTERPRO: IPR002360; -
DR PFM: PF00904; INVOLUCRIN; 42.
DR PROSITE: PS00795; INVOLUCRIN; 1.
DR KeraLinocyte; Repeat.
KW SEQUENCE 544 AA; 63927 MW; 2A02ABA5E1499F9D CRC64;
SQ
Query Match 18.9%; Score 71; DB 1; Length 544;
Best Local Similarity 25.6%; Pred. No. 2.6;
Matches 21; Conservative 17; Mismatches 24; Indels 20; Gaps 2;
OY 5 QQRVEDCDRHCEQDEPRQY-----OCORCQDEQRRHGKRGDL----- 44
DB 354 QDEEGGVKHLDEQDEQLKLEQEGQPKHLEQLKLEHLEQDEGQLKLEQREQL 413
OY 45 MNPORGSGRYEEGEEKSDNP 66
DB 414 LPEQVGQSKHLEFEKQLHR 435

```

```

RT "The human androgen receptor: complementary deoxyribonucleic acid
RT cloning, sequence analysis and gene expression in prostate.";
RL Mol. Endocrinol. 2:1265-1275(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90083302; PubMed=2594783;
RA Lubahn D.B., Brown T.R., Simental J.A., Higgs H.N., Migeon C.J.,
RA Wilson E.M., French F.S.;
RT "Sequence of the intron/exon junctions of the coding region of the
RT human androgen receptor gene and identification of a point mutation
RT in a family with complete androgen insensitivity.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9534-9538(1989).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=90258935; PubMed=2342476;
RA Govindan M.V.;
RT "Specific region in hormone binding domain is essential for hormone
RT binding and trans-activation by human androgen receptor.";
RL Mol. Endocrinol. 4:417-427(1990).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=89017166; PubMed=3174628;
RA Chang C., Kokontis J., Liao S.;
RT "Structural analysis of complementary DNA and amino acid sequences of
RT human and rat androgen receptors.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:7211-7215(1988).
RN [5]
RP SEQUENCE FROM N.A.
RX TISSUE=PROSTATE;
RC MEDLINE=89098909; PubMed=2911578;
RA Tilley W.D., Marcelli M., Wilson J.D., McPhaul M.J.;
RT "Characterization and expression of a cDNA encoding the human androgen
RT receptor.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:327-331(1989).
RN [6]
RP SEQUENCE FROM N.A.
RX TISSUE=PROSTATE;
RC MEDLINE=91155943; PubMed=2293020;
RA Marcelli M., Tilley W.D., Wilson C.M., Griffin J.E., Wilson J.D.,
RA McPhaul M.J.;
RT "Definition of the human androgen receptor gene structure permits the
RT identification of mutations that cause androgen resistance: premature
RT termination of the receptor protein at amino acid residue 588 causes
RT complete androgen resistance.";
RL Mol. Endocrinol. 4:1105-1116(1990).
RN [7]
RP SEQUENCE OF 189-919 FROM N.A.
RX MEDLINE=88178111; PubMed=3353726;
RA Chang C., Kokontis J., Liao S.;
RT "Molecular cloning of human and rat complementary DNA encoding
RT androgen receptors.";
RL Science 240:324-326(1988).
RN [8]
RP SEQUENCE OF 468-919 FROM N.A.
RX MEDLINE=88240407; PubMed=3377788;
RA Tripan J., Klaassen P., Kuiper G.G.J.M., van der Korput J.A.G.M.,
RA Faber P.W., van Rooij H.C.J., Geurts van Kessel A., Voorhorst M.M.,
RA Mulder E., Brinkmann A.O.;
RT "Cloning, structure and expression of a cDNA encoding the human
RT androgen receptor.";
RL Biochem. Biophys. Res. Commun. 153:241-248(1988).
RN [9]
RP POLYMORPHISM OF POLY-GLN REGION.
RX MEDLINE=92220629; PubMed=1561105;
RA Sledzews H.F., Oostra B.A., Brinkmann A.O., Tripan J.;
RT "Tritnucleotide repeat polymorphism in the androgen receptor gene
RT (AR).";
RL Nucleic Acids Res. 20:1427-1427(1992).
RN [10]
RP POLYMORPHISM OF POLY-GLY REGION.
RC TISSUE=BLOOD;
RA Lu J., Danielsen M.;

```

RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 RN [11]
 RP VARIANTS SBWA IN POLY-GLN REGION.
 RX MEDLINE=91287825; PubMed=2062380;
 RA la Spada A.R., Wilson E.M., Lubahn D.B., Harding A.E., Fischback K.H.;
 RT "Androgen receptor gene mutations in X-linked spinal and bulbar
 RT muscular atrophy."
 RL Nature 352:77-79(1991).
 RN [12]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=95023089; PubMed=7937057;
 RA Patterson M.N., Hughes L.A., Gottlieb B., Pinsky L.;
 RT "The androgen receptor gene mutations database."
 RL Nucleic Acids Res. 22:3560-3562(1994).
 RN [13]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=97169385; PubMed=9016528;
 RA Gottlieb B., Trifiro M., Lumbroso R., Vasilion D.M., Pinsky L.;
 RT "The androgen receptor gene mutations database."
 RL Nucleic Acids Res. 25:158-162(1997).
 RN [14]
 RP VARIANT LNCAP ALA-877.
 RX MEDLINE=91083633; PubMed=2260966;
 RA Veldechoite J., Ris-Stalpers C., Kuiper G.G., Jenster G.,
 RA Berrevoets C., Claassen E., van Rooij H.C.J., Trapman J.,
 RA Brinkmann A.O., Mulder E.;
 RT "A mutation in the ligand binding domain of the androgen receptor of
 RT human LNCap cells affects steroid binding characteristics and
 RT response to anti-androgens."
 RL Biochem. Biophys. Res. Commun. 173:534-540(1990).
 RN [15]
 RP VARIANT CAIS MET-866.
 RX MEDLINE=91186983; PubMed=2082179;
 RA Brown T.R., Lubahn D.B., Wilson E.M., French F.S., Migeon C.J.,
 RA Corfen J.L.;
 RT "Functional characterization of naturally occurring mutant androgen
 RT receptors from subjects with complete androgen insensitivity."
 RL Mol. Endocrinol. 4:1759-1772(1990).
 RN [16]
 RP VARIANT CYS-774.
 RX MEDLINE=91310758; PubMed=1856263;
 RA Marcello M., Tilley W.D., Zoppi S., Griffin J.E., Wilson J.D.,
 RA McPhaul M.J.;
 RT "Androgen resistance associated with a mutation of the androgen
 RT receptor at amino acid 772 (Arg-->Cys) results from a combination of
 RT decreased messenger ribonucleic acid levels and impairment of
 RT receptor function."
 RL J. Clin. Endocrinol. Metab. 73:318-325(1991).
 RN [17]
 RP VARIANT CAIS ASN-695 AND HIS-695, AND SEQUENCE OF 629-723 FROM N.A.
 RX MEDLINE=92131007; PubMed=1775137;
 RA Ris-Stalpers C., Trifiro M.A., Kuiper G.G., Jenster G., Romalo G.,
 RA Sai T., van Rooij H.C., Kaufman M., Rosenfield R.L., Liao S.;
 RT "Substitution of aspartic acid-686 by histidine or asparagine in the
 RT human androgen receptor leads to a functionally inactive protein with
 RT altered hormone-binding characteristics."
 RL Mol. Endocrinol. 5:1562-1569(1991).
 RN [18]
 RP VARIANTS CAIS AND PATS.
 RX MEDLINE=93338440; PubMed=1307250;
 RA Batch J.A., Williams D.M., Davies H.R., Brown B.D., Evans B.A.J.,
 RA Hughes I.A., Patterson M.N.;
 RT "Androgen receptor gene mutations identified by SSCP in fourteen
 RT subjects with androgen insensitivity syndrome."
 RL Hum. Mol. Genet. 1:497-503(1992).
 RN [19]
 RP VARIANT CAIS VAL-787.
 RX MEDLINE=92235226; PubMed=1569163;
 RA Nakao R., Haji M., Yanase T., Ogo A., Takayanagi R., Katsube T.,
 RA Fukumaki Y., Nawata H.;
 RT "A single amino acid substitution (Met-786-->Val) in the steroid-
 RT binding domain of human androgen receptor leads to complete androgen
 RT insensitivity syndrome."

RL J. Clin. Endocrinol. Metab. 74:1152-1157(1992).
 RN [20]
 RP VARIANT LNCAP ALA-877.
 RX MEDLINE=92222955; PubMed=1562539;
 RA Veldechoite J., Berrevoets C.A., Ris-Stalpers C., Kuiper G.G.,
 RA Jenster G., Trapman J., Brinkmann A.O., Mulder E.;
 RT "The androgen receptor in LNCap cells contains a mutation in the
 RT ligand binding domain which affects steroid binding characteristics
 RT and response to antiandrogens."
 RL J. Steroid Biochem. Mol. Biol. 41:665-669(1992).
 RN [21]
 RP VARIANT MET-730.
 RX MEDLINE=92335289; PubMed=1631125;
 RA Newmark J.R., Hardy D.O., Tonb D.C., Carter B.S., Epstein J.L.,
 RA Isaacs W.B., Brown T.R., Barrack E.R.;
 RT "Androgen receptor gene mutations in human prostate cancer."
 RL Proc. Natl. Acad. Sci. U.S.A. 89:6319-6323(1992).
 RN [22]
 RP VARIANT CAIS VAL-754.
 RX MEDLINE=93372806; PubMed=8103398;
 RA Iobaccaro J.-M., Lumbroso S., Ktari R., Dumas R., Sultan C.;
 RT "An exonic point mutation creates a Maelfi site in the androgen
 RT receptor gene of a family with complete androgen insensitivity
 RT syndrome."
 RL Hum. Mol. Genet. 2:1041-1043(1993).
 RN [23]
 RP VARIANT CAIS ARG-807.
 RX MEDLINE=94108430; PubMed=8281140;
 RA Adeyemo O., Kallilo P.J., Palvimo J.J., Kontula K., Jaenke O.A.;
 RT "A single-base substitution in exon 6 of the androgen receptor gene
 RT causing complete androgen insensitivity: the mutated receptor fails
 RT to transactivate but binds to DNA in vitro."
 RL Hum. Mol. Genet. 2:1809-1812(1993).
 RN [24]
 RP VARIANT PATS VAL-743.
 RX MEDLINE=93315568; PubMed=8325932;
 RA Nakao R., Yanase T., Sakai Y., Haji M., Nawata H.;
 RT "A single amino acid substitution (Gly743-->Val) in the steroid-
 RT binding domain of the human androgen receptor leads to Reifenstein

Query Match 18.9%; Score 71; DB 1; Length 919;
 Best Local Similarity 29.0%; Pred. No. 4.2;
 Matches 18; Conservative 14; Mismatches 28; Indels 2; Gaps 1;
 QY 5 GOREVEDCRHEDDEPRLOYOCORRCOROHGSGGLMPOROGSGRYEGEKGSD 64
 DB 58 QQQ
 QY 65 NP 66
 DB 116 QP 117

RESULT 12
 TAGB_DICDI STANDARD; PRT; 1905 AA.
 ID TAGB_DICDI
 AC P54683;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE PRESTALK-SPECIFIC PROTEIN TAGB PRECURSOR (EC 3.4.21.-).
 GN TAGB.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Dictyostelida; Dictyostelium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX4;
 RX MEDLINE=95262903; PubMed=7744252;
 RA Shaunsky G., Kuspa A., Loomis W.F.;
 RT "A multidrug resistance transporter/serine protease gene is required
 RT for prestalk specialization in Dictyostelium."
 RL Genes Dev. 9:1111-1122(1995).
 CC -i- FUNCTION: INTERCELLULAR COMMUNICATION VIA TAGB MAY MEDIATE

[illegible]

OY 4 POOREYEDCRRHCBOEPRLQYOCORRCOPOBOPOHGRGGDLNMPOR 49
111: : 11:1:1 11: : 111: 111:
DB 80 POOQPOMO-OEOWPQOQPOVVOOQOQWPQOQHRRQHGQOQOQCMNSQO 124

Search completed: March 1, 2001, 16:16:50
Job time: 430 sec